AD-A085 831 BROWN UNIV PROVIDENCE RI F/G 12/1

AN AVERAGING METHOD FOR STOCHASTIC APPROXIMATIONS WITH CONSTANT--ETC(U)

MAR 80 H J KUSHNER N00014-76-C-0279

UNCLASSIFIED NL

1 OF 1

AD A
085831

END

DATE
FILMED

8-80

DTIC

ADA085831

4 12

6 AN AVERAGING METHOD FOR STOCHASTIC APPROXIMATIONS

WITH CONSTANT PARAMETERS: SMALL PARAMETER VALUES

16 H. J. Kushner*

Abstract

Stochastic approximations with constant gain coefficients and dependent noise and nonlinear or even discontinuous dynamics have many applications in control, automata and communication theory. When the gain coefficient is small, an asymptotic theory is developed which gives much information on the character of the paths and errors. The method involves both "averaging" and "stability" ideas. The ideas are outlined. An example which illustrates the basic ideas and techniques is given.

11 Mar [redacted]80

12 24

DTIC
JUN 1980
A

15 N00014-76-C-0279,
AFOSR 77-3063

*Brown University, Divisions of Applied Mathematics and Engineering. This research ~~was supported by~~ the Office of Naval Research (N00014-76-C-0279-~~P0002~~), the Air Force Office of Scientific Research (AFOSR 77-3063), and the National Science Foundation (NSF-Eng. 77-12946).

DDC FILE COPY

065258 80 5 19 094

I. Introduction

Several quite powerful methods are available for analyzing asymptotic properties of many kinds of stochastic approximations with gain sequences which are either constants, or tend to zero. Methods for the first type of gain sequence - examples of which abound - are much less well known. In this paper, we develop a method for such a problem, based on "stability", "averaging" and "diffusion approximations via weak convergence theory" - ideas which have served very well in many other areas. The techniques have great power and applicability. The general idea will be outlined and a simple example dealt with in detail. Despite the simplicity of the example, it well illustrates the general approach, the kind of calculations which need to be done and the type of results which can be expected. The example is used simply as a vehicle for explaining the main ideas. Nevertheless, only the surface of a large subject will be touched. There is much work on the problem (when the gain sequence tends to zero) of the asymptotic properties of stochastic approximation via the study of the stability of an associated ordinary differential equation [1], [2]. The local asymptotic behavior (near the limit points), on the other hand, is obtained via the study of an associated stochastic differential equation (as in rate of convergence studies [1], [3], [4]). Here similar intentions are pursued for process with constant (but small) gains, state-dependent noise, and perhaps discontinuous forcing functions.

We are concerned with asymptotic properties ($n \to \infty$, then $\varepsilon \to 0$) of a subclass of vector stochastic difference equations of the form

Accession For

(1.1) $$Y_{n+1}^{\varepsilon} = Y_n^{\varepsilon} + \varepsilon h_{\varepsilon}(Y_n^{\varepsilon},\xi_n^{\varepsilon}) + \sqrt{\varepsilon}\, g_{\varepsilon}(Y_n^{\varepsilon},\xi_n^{\varepsilon}) + o(\varepsilon),$$

$Y_n^{\varepsilon} \in R^r$ = Euclidean r-space,

A

where h_ε and g_ε are not necessarily continuous everywhere and $\{\xi_n^\varepsilon\}$ is a sequence of correlated (and perhaps $\{Y_n^\varepsilon\}$-dependent) random variables, for each $\varepsilon > 0$. Equations such as (1.1) occur very frequently in applications in stochastic control, communication and automata theory. Often the g_ε or h_ε are indicator functions; i.e., the iterate Y_n^ε moves "left" or "right" by ε, depending only on whether some event or other occurred.

There is a vast (stochastic approximation) literature on (1.1) when ε is replaced by a sequence $\varepsilon_n \to 0$. But, very frequently in implementations of stochastic approximation, ε_n is either held constant or else $\varepsilon_n \downarrow \varepsilon > 0$ as $n \to \infty$, owing to the desire to track changes or for the purpose of improving robustness. Consider a particular case of (1.1), a scalar Robbins-Monro procedure of the form

$$(1.2) \qquad X_{n+1}^\varepsilon = X_n^\varepsilon - \varepsilon \operatorname{sign}[k(X_n^\varepsilon)+\xi_n^\varepsilon],$$

where ξ_n^ε has a symmetric distribution, $\{\xi_n^\varepsilon\}$ are not necessarily independent (and might even depend on the iterates), $k(\cdot)$ is continuous, and there is a θ such that $k(x) > 0$ for $x > \theta$, and $k(x) < 0$ for $x < \theta$. In particular, suppose that for $x \geq 0$ (resp., $x \leq 0$), $k(x)$ is bounded below (above, resp.) by an increasing function which has a non-zero slope near the origin. We might want to prove that $X_n^\varepsilon \xrightarrow{P} \theta$ as $n \to \infty$ and $\varepsilon \to 0$, or even to get a good idea of the statistical structure of the tail of the sequence $\{X_n^\varepsilon - \theta\}$. Define $U_n^\varepsilon = (X_n^\varepsilon - \theta)/\sqrt{\varepsilon}$. Something akin to a rate of convergence can be obtained by studying the asymptotic properties of $\{U_n^\varepsilon\}$ - which, as we'll see, brings us back to a process of the form (1.1) with a $\sqrt{\varepsilon}$ term included.

There are numerous other applications of (1.1) - particularly to systems whose dynamics are determined by "logical" criteria - where the movement is determined partly by the satisfaction of a "logical" criterion. Such systems have not received much attention, important though they are. One very clever approach and some nice applications appear in [5], [6]. Although that method and ours both exploit "averaging" phenomena, our method is different. We are not confined to finite intervals [n: $\varepsilon n \leq T$] for some arbitrary T (indeed, it is the "tail" which is of most interest here), and [5] and [6] did not explicitly deal with a $\sqrt{\varepsilon}$ factor in (1.1). The $\sqrt{\varepsilon}$ factor always occurs when a local analysis (involving U_n^ε) is done. When this factor is present the limiting determining equation is a stochastic rather than ordinary differential equation. We concentrate on the asymptotic part of $\{X_n^\varepsilon\}$ or $\{U_n^\varepsilon\}$, for small ε, often the part of greatest interest. One rather direct but useful method for the asymptotic problem in in [4].

In practice ε is often small. But, generally, it is very hard to get information on what happens when ε is not small, just as in stochastic approximation it is hard to get information on the theoretical behavior when n is not large. A main question is how well the theory for small ε predicts what happens in other cases. Simulations on continuous parameter problems which resemble these in certain respects indicate that the prediction is good for many cases when the parameter ε is in a "normal" range. In some cases, the asymptotic behavior is better than predicted. Generally, the closeness of the prediction to the behavior actually observed seems to depend heavily on the form of the nonlinearities, and on the correlation structure of the noise, and generalizations are hard to make. By concentrating on large n, we are effectively concentrating on what happens after the "transient" period is over.

The techniques used here come from references [7], [8], [11], which concern the general problem (1.1) (or continuous time analogues) under various sets of assumptions. The emphasis in [7] is on the case where g_ε, h_ε are smooth, but an outline is given of the method for the more general case. Reference [8] treats two problems in great detail, one arising in an automata approach to telephone traffic routing, and the other a recursive quantizer in communication theory. Our purpose here is to describe the general method, citing results from [7], [8] and avoiding duplication of proofs wherever convenient. As an illustration of the power and usefulness of the method, a detailed analysis of (1.2) (for both the tail of $\{X_n^\varepsilon\}$ and $\{U_n^\varepsilon\}$) will be given. The development will illustrate the usefulness of the approach for other problems. The case (1.2) is simple (but not trivial), but a very similar method of analysis is used in more general cases.

Outline of the paper. The first important result (Theorem 2) concerns the tightness of $\{U_n^\varepsilon$, large n, small $\varepsilon\}$. By this we mean that there is an $\varepsilon_0 > 0$, whose value is not important, such that for each $\varepsilon \le \varepsilon_0$, there is an integer $N_\varepsilon < \infty$ such that

$$(1.3)\qquad \lim_{K\to\infty} \sup_{\substack{\varepsilon\le\varepsilon_0 \\ n\ge N_\varepsilon}} P\{|U_n^\varepsilon| \ge K\} = 0.$$

I.e., $\{U_n^\varepsilon$, large n, small $\varepsilon\}$ is bounded in probability. Such a result makes possible a detailed asymptotic analysis of $\{U_n^\varepsilon\}$, since it implies that the "tails" of $\{X_n^\varepsilon\}$ are uniformly close to θ in a specific statistical sense. To simplify the development most details are for the special case (1.2). More general cases can readily be dealt with in a very similar way, at the expense of a somewhat heavier notation, but the simple case illustrates the main ideas and methods.

Next, we define a continuous parameter process $U^\varepsilon(\cdot)$. The $\{N_\varepsilon\}$ will always satisfy (1.3). Let $\{n_\varepsilon\}$ denote any sequence of integers satisfying $n_\varepsilon \ge N_\varepsilon$ and $n_\varepsilon \to \infty$ as $\varepsilon \to 0$. Their values will either be stated when needed or will be unimportant. Define the process $U^\varepsilon(t) = U^\varepsilon_{n_\varepsilon+i}$ on $[i\varepsilon, i\varepsilon+\varepsilon)$. Define $Y^\varepsilon(\cdot)$ by $Y^\varepsilon(0) = Y^\varepsilon_0$ and $Y^\varepsilon(t) = Y^\varepsilon_n$ on $[\varepsilon n, \varepsilon n+\varepsilon)$ (refer to (1.1)). The general result is that $U^\varepsilon(\cdot)$ converges (under suitable conditions) weakly in $D^r[0,\infty)$ to a Gauss-Markov diffusion $U(\cdot)$ which satisfies an equation of the form

$$(1.4) \qquad dU = -GUdt + \sigma dB,$$

where $B(\cdot)$ is a standard Browninan motion and $G > 0$. If $n_\varepsilon \to \infty$ fast enough as $\varepsilon \to 0$, then the limit $U(\cdot)$ is stationary. <u>All terms, conditions and parameters will be defined below</u>, but first, let us examine the limit (1.4).

For small ε, the processes $\{X^\varepsilon_n, U^\varepsilon_n\}$ (or $\{Y^\varepsilon_n\}$, for the case (1.1)) have a (perhaps long) transient period, before their distributions "settle down" - especially if $(X^\varepsilon_0-\theta)$ is large compared to ε. We are concerned only with the asymptotic part - after the so-called transient period is over; i.e. when $U^\varepsilon_n \sim O(1)$. We can get the variance of the asymptotic part from (1.4), since $X^\varepsilon_n \sim \sqrt{\varepsilon}\, U^\varepsilon_n+\theta$. Also, (1.4) gives the local correlation structure of the asymptotic part of the process. Weak convergence methods are a very natural and convenient tool for getting our results. The details of the derivation of (1.4) will be given for the model (1.2). Very similar methods can be used for more complex problems [7], [8]. Next, some comments and definitions concerned with weak convergence theory will be given. Then we comment on the so-called martingale problem of Strook and Varadhan [9], which provides a characterization of the desired limit process which is convenient from the point of view of simplifying the proof of showing that it actually is the limit. Section II contains the main background theorems. For (1.4), tightness of $\{U^\varepsilon_n$, large n, small $\varepsilon\}$ is proved in Section III, and in Section IV we show how to get (1.4) for the case of model (1.2).

A note on weak convergence theory. Only a few comments will be made. For full details see [10], or [1], Chapter 3, for a brief summary. $D^r[0,\infty)$ denotes the space of R^r-valued functions on $[0,\infty)$ which are right continuous and have left-hand limits. The "process" $U^\varepsilon(\cdot)$ is treated as a random variable defined on the sample space $D^r[0,\infty)$ and induces a measure on it which we denote by P_ε (a useful topology, called the "Skorokhod" topology, is used on D^r; see [10]). $\{U^\varepsilon(\cdot)\}$ is said to be tight iff for each $\delta > 0$ there is a compact set $K_\delta \in D^r[0,\infty)$ such that $P_\varepsilon\{K_\delta\} \geq 1-\delta$, all ε. $U^\varepsilon(\cdot)$ is said to converge weakly to $U(\cdot)$ if $U(\cdot)$ has paths in $D^r[0,\infty)$, and induces a measure P on it, and for each real-valued continuous function $F(\cdot)$ on $D^r[0,\infty)$, $\int F(v)dP_\varepsilon(v) \to \int F(v)dP(v)$ as $\varepsilon \to 0$. The basic result is: If $\{U^\varepsilon(\cdot)\}$ is tight on $D^r[0,\infty)$, then each subsequence contains a further subsequence which converges weakly to some process with paths in $D^r[0,\infty)$. Our job here is to characterize the limit process and to show that it does not depend on the subsequence. Thus weak convergence is a substantial extension of convergence in distribution. It is a tool that is extremely useful in many areas of applied probability where limit or approximation problems are of concern. Criteria for tightness and weak convergence are often given in terms of the multivariate distributions of the processes $\{U^\varepsilon(\cdot)\}$. See Section II.

A note on the martingale problem. This problem arises because in the weak convergence analysis it is convenient to characterize the limit $U(\cdot)$, whatever it may be, by showing that it solves a certain set of equations which are known as the martingale problem. Then we show that the $U(\cdot)$ of (1.4) is the only solution to that martingale problem. Let $x(\cdot)$ be the solution to the stochastic differential equation (SDE)

(1.5) $\quad dx = b(x,t)dt + \sigma(x,t)dB, \quad B(\cdot) = \text{standard Brownian motion in } R^r,$

and set $\sigma(x)\sigma'(x)/2 = a(x)$. Suppose that $b(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ are continuous in (x,t) and satisfy a uniform Lipschitz and linear growth condition (in x). Then it is well known that the SDE (1.5) has a unique solution in the Itô sense. Since $\int_0^t \sigma(x_s)dB_s$ is a martingale, for any smooth bounded function $f(\cdot)$, Itô's lemma implies that

$$(1.6)\qquad f(x(t),t) - f(x(0),0) - \int_0^t (A + \frac{\partial}{\partial s}) f(x(s),s)ds \equiv M_f(t)$$

is a martingale for each initial condition $x(0) = x$, where

$$(1.7)\qquad A = \sum_i b_i(x,t) \frac{\partial}{\partial x_i} + \sum_{i,j} a_{ij}(x,t) \frac{\partial^2}{\partial x_i \partial x_j}$$

$$= \text{differential generator of (1.5).}$$

Let $y(\cdot)$ denote the generic element of $D^r[0,\infty)$, let $\hat{\mathscr{C}}_0$ denote the continuous functions on $R^r \times [0,\infty)$ with compact support and let $\hat{\mathscr{C}}_0^{\alpha,\beta}$ denote the subclass whose mixed α t-derivatives and β x-derivatives are continuous. For $f \in \hat{\mathscr{C}}_0^{1,2}$, define

$$(1.8)\qquad \overline{M}_f(t) = f(y(t),t) - f(y(0),0) - \int_0^t (A + \frac{\partial}{\partial s}) f(y(s),s)ds,$$

where A is an operator of the form (1.7) - but not necessarily satisfying the Lipschitz and growth conditions. If for each x, there is a measure P_x on $D^r[0,\infty)$ such that $P_x\{y(0)=x\} = 1$ and $\overline{M}_f(\cdot)$ is a P_x-martingale for each $f \in \hat{\mathscr{C}}_0^{1,2}$, then $\{P_x\}$ is said to solve the martingale problem (of Strook and Varadhan [9]). If P_x is unique for each x (as it is under the Lipschitz and growth condition cited above), then the $\{P_x\}$ induce a Markov process on the sample space $D^r[0,\infty)$. Also, for each x, there is a $B(\cdot)$ such that under P_x, the process satisfies (1.5) with initial condition x.

The martingale formulation of SDEs is very useful in convergence studies. The basic background theorems for our case [7], [11] are proved by first showing tightness of $\{U^\varepsilon(\cdot)\}$ and then showing that $\overline{M}_f(\cdot)$ is a martingale for each $f \in \hat{\mathscr{C}}_0^{2,3}$ when any weak limit $\tilde{U}(\cdot)$ is substituted for $y(\cdot)$ and P_x is replaced by the measure of $\tilde{U}(\cdot)$ and A is the operator (1.7) (in our case this specializes to the operator of (1.4)). Below, an assumption concerning uniqueness of the solution of the martingale problem will be made. This is needed, for otherwise the weak limits of $\{U^\varepsilon(\cdot)\}$ might not be unique.

<u>Truncation</u>. For mathematical as well as practical reasons, it is helpful to work with a <u>truncated</u> $\{X_n^\varepsilon\}$. The particular technical difficulty introduced in the untruncated case will be discussed after the proof of Theorem 2. The truncation introduced now alters the basic problem (1.2) and has nothing to do with the "technical" truncation $\{X_n^{\varepsilon,N}\}$ introduced in the first paragraph of Section II. Suppose that we know real numbers x_ℓ, x_u such that $\theta \in (x_\ell, x_u)$. Let the bar $|$ denote truncation. Then we replace (1.2) by

$$(1.9) \qquad X_{n+1}^\varepsilon = (X_n^\varepsilon - \varepsilon \operatorname{sign}[k(X_n^\varepsilon)+\xi_n^\varepsilon])\Big|_{x_\ell}^{x_u}.$$

<u>Equation</u> (1.9) <u>is the actual algorithm whose asymptotic properties are to be studied</u>. There are continuous functions $a_\varepsilon(\cdot)$, $b_\varepsilon(\cdot)$ such that $a_\varepsilon(x) = b_\varepsilon(x) = \varepsilon$ in $[x_\ell+\varepsilon, x_u-\varepsilon]$ and both functions have values in $[0,\varepsilon]$ and are infinitely differentiable functions of x except possibly at the points $x_\ell+\varepsilon$, $x_u-\varepsilon$, and they are such that

(1.10) $X_{n+1}^{\varepsilon} = X_n^{\varepsilon} - a_{\varepsilon}(X_n^{\varepsilon})\ I\{k(X_n^{\varepsilon})+\xi_n^{\varepsilon}\geq 0\} + b_{\varepsilon}(X_n^{\varepsilon})\ I\{k(X_n^{\varepsilon})+\xi_n^{\varepsilon}<0\}$,

where $I\{\cdot\}$ is the indicator function.

II. Weak Convergence Theorems

Notation and a comment on tightness. The theorem below, taken from [7], assumes tightness of $\{Y^{\varepsilon}(\cdot)\}$ (or $\{U^{\varepsilon}(\cdot)\}$). Such tightness is not hard to get under reasonable conditions by Theorem 2 of [7] if $\{Y^{\varepsilon}(\cdot)\}$ (or $\{U^{\varepsilon}(\cdot)\}$) is bounded. This will be further commented on below. We can bound the processes by a truncation device which, since it is used only as a technical tool, does not lose us any generality, and it will now be explained. Loosely speaking, if the truncated processes defined below exhibit a suitable weak convergence, then so do the $\{U^{\varepsilon}(\cdot)\}$ and $\{Y^{\varepsilon}(\cdot)\}$ as originally defined. Define $S_N = \{x: |x|\leq N\}$, and let $b_N(x)$ be a function with values in [0,1], equal to unity in S_N, equal to zero out of S_{N+1} and infinitely differentiable. Define $Y_n^{\varepsilon,N}$ (and for (1.2), $U_n^{\varepsilon,N}$) by

$$(2.1)\qquad Y_{n+1}^{\varepsilon,N} = Y_n^{\varepsilon,N} + [\varepsilon h_{\varepsilon}(Y_n^{\varepsilon,N},\xi_n^{\varepsilon}) + \sqrt{\varepsilon}\ g_{\varepsilon}(Y_n^{\varepsilon,N},\xi_n^{\varepsilon}) + o(\varepsilon)]b_N(Y_n^{\varepsilon,N}),$$

$$(2.2)\qquad U_{n+1}^{\varepsilon,N} = U_n^{\varepsilon,N} - \sqrt{\varepsilon}\ \mathrm{sign}[k(X_n^{\varepsilon,N})+\xi_n^{\varepsilon}]b_N(U_n^{\varepsilon,N}),\qquad X_n^{\varepsilon,N} = \sqrt{\varepsilon}\ U_n^{\varepsilon,N} + \theta.$$

Let $Y^{\varepsilon,N}(\cdot)$ and $U^{\varepsilon,N}(\cdot)$ denote the corresponding interpolations of $\{Y_n^{\varepsilon,N}\}$ and $\{U_n^{\varepsilon,N}\}$. We use $U^{\varepsilon,N}(0) = U_{N_\varepsilon}^{\varepsilon}$ if $|U_{N_\varepsilon}^{\varepsilon}| \leq N$, and set it equal to zero otherwise. Similarly for $Y^{\varepsilon,N}(0)$. Owing to the tightness of $\{U_n^{\varepsilon}\}$, this causes no problem.

We will see below that (2.2) is consistent with (1.9) - (1.10) in that the weak limits of $U^{\varepsilon}(\cdot)$, where $U^{\varepsilon}(\cdot)$ is obtained from (1.9) - (1.10), are the same as those which Theorem 1 below gives us for (2.2), as $\varepsilon \to 0$, then $N \to \infty$.

Suppose for the moment that for each N, $\{Y^{\varepsilon,N}(\cdot)\}$ or $\{U^{\varepsilon,N}(\cdot)\}$ converges weakly to $U^N(\cdot)$, a diffusion process (with differential operator denoted by A^N) whose part up to first escape from S_N equals the part of $U(\cdot)$ up to first escape from S_N. Then if the solution to the martingale problem for operator A (A being the differential generator of $U(\cdot)$) is unique, $U^{\varepsilon}(\cdot) \to U(\cdot)$ weakly. Let $E_n^{\varepsilon,N}$ denote expectation conditioned on $\{Y_i^{\varepsilon,N}, i\le n, \xi_i^{\varepsilon}, i<n\}$ or on $\{X_i^{\varepsilon,N}, i\le n, \xi_i^{\varepsilon}, i<n\}$, depending on the case (2.1) or (2.2). E_n^{ε} is similarly defined when the superscript N is absent.

The following is an adaptation of Theorem 1 of [7] to our case. It is stated for the general case (1.1), (2.1).

Theorem 1. Let the differential generators A and A^N have continuous coefficients, which are equal in S_N for each N. Let the solution to the martingale problem corresponding to operator A be unique on $D^r[0,\infty)$, for each $x \in R^r$. For each N and $f(\cdot,\cdot) \in \mathscr{D}$, a dense set (sup norm) in $\hat{\mathscr{C}}_0$, let there be a sequence $\{f^{\varepsilon,N}(\cdot)\}$ of random functions satisfying the following conditions. Each is constant on each $[n\varepsilon, n\varepsilon+\varepsilon)$ interval, at $n\varepsilon$ it is measurable with respect to the σ-algebra induced by $\{Y_j^{\varepsilon,N}, j\le n, \xi_j^{\varepsilon}, j<n\}$ and for each N and t

$$(2.3)\qquad \sup_{n,\varepsilon} E|f^{\varepsilon,N}(\varepsilon n)| + \sup_{n,\varepsilon} \frac{1}{\varepsilon} E|E_n^{\varepsilon,N} f^{\varepsilon,N}(\varepsilon n+\varepsilon) - f^{\varepsilon,N}(\varepsilon n)| < \infty,$$

$$(2.4)\qquad E|f^{\varepsilon,N}(\varepsilon n) - f(Y_n^{\varepsilon,N},\varepsilon n)| \to 0 \qquad \text{as } \varepsilon \to 0 \text{ and } \varepsilon n = t,$$

$$(2.5)\qquad E\left|\frac{E_n^{\varepsilon,N} f^{\varepsilon,N}(\varepsilon n+\varepsilon) - f^{\varepsilon,N}(\varepsilon n)}{\varepsilon} - \left(\frac{\partial}{\partial t} + A^N\right) f(Y_n^{\varepsilon,N},\varepsilon n)\right| \to 0 \quad \text{as } \varepsilon \to 0 \text{ and } \varepsilon n = t.$$

Then if $\{Y^{\varepsilon,N}(\cdot), 0<\varepsilon\le\varepsilon_0\}$ is tight on $D^r[0,\infty)$ for each N, where ε_0 doesn't depend on N, and $Y^\varepsilon(0)$ converges in distribution to $U(0)$ as $\varepsilon \to 0$, we have $Y^\varepsilon(\cdot) \to U(\cdot)$ weakly, the unique solution to the martingale problem for operator A with initial condition $U(0)$.

The main burden of proof is in finding the functions $\{f^{\varepsilon,N}(\cdot)\}$, and the method for this will be developed in Theorem 3. In our case, the $\{U^{\varepsilon,N}(\cdot)\}$ can be shown to be tight via the method of [7], Theorem 2, which also makes use of the functions $\{f^{\varepsilon,N}(\cdot)\}$.

Define the operator $\hat{A}^{\varepsilon,N}$ by $\hat{A}^{\varepsilon,N} f^{\varepsilon,N}(\varepsilon n) = \frac{1}{\varepsilon} E_n^{\varepsilon,N}[f^{\varepsilon,N}(\varepsilon n+\varepsilon) - f^{\varepsilon,N}(\varepsilon n)]$. $\hat{A}^{\varepsilon,N}$ has the character of an infinitesimal operator. Considering the $f \in \mathscr{D}$ as "test functions", the aim of the theorem is to find functions $f^{\varepsilon,N}$ which are close to the test functions and such that the action of $\hat{A}^{\varepsilon,N}$ on $f^{\varepsilon,N}$ is close to the action of $(A^N + \frac{\partial}{\partial t})$ on the test function f, for each N.

III. Tightness of $\{U_n^\varepsilon$, small ε, large n$\}$

Assumptions. Henceforth we stick to problem (1.9) or (1.10). Since we treat the case where the $\{\xi_n^\varepsilon\}$ can depend on the $\{X_n^\varepsilon\}$, some information on the nature of the dependence must be provided. If $\{\xi_n^\varepsilon\}$ were a sequence of independent random variables, then the assumptions and development would be much simpler. But it is worthwhile to do the general case, since various forms of it occur frequently. We assume that there is an auxiliary process $\{\hat{\xi}_n^\varepsilon\}$ such that $\{\hat{\xi}_{n-1}^\varepsilon, X_n^\varepsilon\}$ is a Markov process with a homogeneous transition function, and that ξ_n^ε is a function of $\hat{\xi}_n^\varepsilon$, X_n^ε. If X_n^ε were held fixed at a value x for all time, then we have a Markov process which we denote by $\{\hat{\xi}_n^\varepsilon(x)\}$ and whose transition probabilities are defined by the marginals of that of $\{\hat{\xi}_{n-1}^\varepsilon, X_n^\varepsilon\}$; i.e., given by $P\{\hat{\xi}_n(x) \in C | \hat{\xi}_{n-1}(x), x\} = P\{\hat{\xi}_n^\varepsilon \in C | \hat{\xi}_{n-1}^\varepsilon = \hat{\xi}_{n-1}(x), X_{n-1}^\varepsilon = x\}$. Now a function $\xi_n(x)$

can be defined, where $\xi_n(x)$ is obtained from $(\hat{\xi}_n(x),x)$ in the same way that ξ_n^ε was obtained from the pair $(\hat{\xi}_n^\varepsilon,X_n^\varepsilon)$. Define the "partial" ℓ-step transition function $P\{\cdot,\cdot,\cdot|\cdot\}$ for $\{\xi_n(x)\}$ by

$$(3.1)\qquad P(\hat{\xi}_j(x),\ell,B|x) = P\{\xi_{j+\ell}(x)\in B|\hat{\xi}_j(x),x\}.$$

The conditions introduced below will be on the rate of convergence of certain ℓ-step transition functions to their limits as $\ell\to\infty$, and on the smoothness of the transition functions. Some smoothness is required in order to "average out" the discontinuity in the sign function. The symbol $K(x)$ will be used to denote the set $[-k(x),\infty)$, and M denotes a constant whose value may change from usage to usage. We use the following conditions, some in Theorem 2 and some in Theorem 3.

(A1) For each x, there is a unique measure $P(\cdot|x)$ such that

$$\sum_{\ell=0}^{\infty} |P(\hat{\xi},\ell,K(x)|x) - P(K(x)|x)| \le M$$

uniformly in $\hat{\xi}$ and in $x\in[x_\ell,x_u]$. $P(\cdot|x)$ would normally be the marginal of the stationary measure of $\{\hat{\xi}_n(x)\}$.

(A2) For each x, $P(\cdot|x)$ has a density $p(\cdot|x)$ which is symmetric about $\xi = 0$. Define $\bar{g}(x) = \int \operatorname{sign}[k(x)+\xi]p(\xi|x)d\xi$. Let $\bar{g}(\theta) = 0$. $\bar{g}(\cdot)$ is differentiable at $x=\theta$ with $\bar{g}_x(\theta) > 0$ and there is a non-negative, non-decreasing function $q(\cdot)$ such that $\bar{g}(x) \ge q(x)$ for $x\ge\theta$ and $\bar{g}(x) \le -q(x)$ for $x\le\theta$ and $q(x)\ne 0$ if $x\ne\theta$. $k(\cdot)$ is continuously differentiable.

(A3) $P([y,\infty)|x)$ and $P(\hat{\xi},\ell,[y,\infty)|x)$ are continuously differentiable in x,y, the continuity being uniform in $\ell>0$, $\hat{\xi}$ and in x,y in bounded sets. Also (the subscripts x, y denote the derivatives)

$$\sum_{\ell=1}^{\infty} |P_x(\hat{\xi},\ell,K(x)|x) - P_x(K(x)|x)| \leq M,$$

uniformly in $\hat{\xi}$ and $x \in [x_\ell, x_u]$.

(A4) $P\{\hat{\xi}_n(x) \in C | \hat{\xi}_{n-1}(x), x\}$ is continuous in x in some neighborhood of θ, uniformly in C and $\hat{\xi}_{n-1}(x)$.

(A5) $P(\hat{\xi},1,[-b,b]|x) \to 0$ as $b \to 0$ uniformly in $\hat{\xi}$ and in x in some neighborhood of θ.

Let P_n^θ denote the regular conditional distribution of $\{\xi_{n+j}(\theta),\ j\geq 0\}$ conditioned on $\hat{\xi}_{n-1}(\theta) = \hat{\xi}_{n-1}^\varepsilon$, the actual sample value at time n-1. Let E_n^θ denote the corresponding regular conditional expectation. We need convergence of the conditional expectations of certain functions. In particular,

(A6) There are functions q(m) which we write as (which define the expectation operator E^θ)

$$q(m) \equiv E^\theta \text{sign}\ \xi_0(\theta)\ \text{sign}\ \xi_m(\theta) \equiv E^\theta\ \text{sign}\ \xi_\ell(\theta)\ \text{sign}\ \xi_{\ell+m}(\theta)$$

such that

$$E_n^\theta\ \text{sign}\ \xi_{i+n}(\theta)\ \text{sign}\ \xi_{i+n+m}(\theta) \to q(m) \qquad \text{as } i \to \infty$$

for all $\hat{\xi}_{n-1}^\varepsilon$ and $n > 0$. Also $\sum_{m=1}^{\infty} |q(m)| < \infty$ and

$$\sum_{j=n}^{\infty} \sum_{\ell=j+1}^{\infty} |E_n^\theta\ \text{sign}\ \xi_j(\theta)\ \text{sign}\ \xi_\ell(\theta) - E^\theta\ \text{sign}\ \xi_j(\theta)\ \text{sign}\ \xi_\ell(\theta)| \leq M$$

for all $\hat{\xi}_{n-1}^\varepsilon$, and $n > 0$.

Define

$$\varepsilon\bar{g}_\varepsilon(x) = \int [a_\varepsilon(x) I\{k(x)+\xi \geq 0\} - b_\varepsilon(x) \; I\{k(x)+\xi<0\}] \; p(\xi|x)d\xi.$$

Note that $\bar{g}_\varepsilon(\cdot)$ has the properties ascribed to $\bar{g}(\cdot)$ in (A2).

In Theorems 2 and 3, all $O(\cdot)$ and $o(\cdot)$ are uniform (in the $O(\cdot)$, $o(\cdot)$ property) in all variables other than their argument.

<u>Theorem 2</u>. <u>Assume</u> (A1) - (A3). <u>There is a sequence of integers</u> $N_\varepsilon \to \infty$ <u>as</u> $\varepsilon \to 0$ <u>and an</u> $\varepsilon_0 > 0$ <u>such that the double sequence</u> $\{U_n^\varepsilon,\ n \geq N_\varepsilon,\ \varepsilon \leq \varepsilon_0\}$ <u>is tight</u>.

<u>Proof</u>. <u>Part 1</u>. The method is based on combination of a Liapunov function and an averaging technique. Define $V(x) = (x-\theta)^2$. Note that

$$\bar{g}_\varepsilon(x) = -a_\varepsilon(x)P(K(x)|x) + b_\varepsilon(x)(1-P(K(x)|x))$$

and

$$E_n^\varepsilon V(X_{n+1}^\varepsilon) - V(X_n^\varepsilon) = V_x(X_n^\varepsilon)E_n^\varepsilon(X_{n+1}^\varepsilon - X_n^\varepsilon) + O(\varepsilon^2), \tag{3.2}$$

$$E_n^\varepsilon(X_{n+1}^\varepsilon - X_n^\varepsilon) = -a_\varepsilon(X_n^\varepsilon)P(\hat{\xi}_{n-1}^\varepsilon,1,K(X_n^\varepsilon)|X_n^\varepsilon) + b_\varepsilon(X_n^\varepsilon)(1-P(\hat{\xi}_{n-1}^\varepsilon,1,K(X_n^\varepsilon)|X_n^\varepsilon)). \tag{3.3}$$

For small ε,

$$E_n^\varepsilon(X_n^\varepsilon-\theta)[-a_\varepsilon(X_n^\varepsilon)I\{k(X_n^\varepsilon)+\xi_n^\varepsilon \geq 0\} + b_\varepsilon(X_n^\varepsilon)I\{k(X_n^\varepsilon)+\xi_n^\varepsilon<0\}]$$

$$\leq \varepsilon E_n^\varepsilon(X_n^\varepsilon-\theta)[-I\{k(X_n^\varepsilon)+\xi_n^\varepsilon \geq 0\} + I\{k(X_n^\varepsilon)+\xi_n^\varepsilon<0\}].$$

Thus

$$E_n^\varepsilon V(X_{n+1}^\varepsilon)-V(X_n^\varepsilon) \le \varepsilon V_x(X_n^\varepsilon)[1 - 2P(\hat{\xi}_{n-1}^\varepsilon,1,K(X_n^\varepsilon)|X_n^\varepsilon)] + O(\varepsilon^2).$$

Now, we introduce a perturbation $V_1^\varepsilon(n)$ to the Liapunov function. It is used as a technical device to allow us to effectively replace the P which appear in (3.3) by the "stationary" probabilities introduced in (A1). Define $V_1^\varepsilon(n)$ by

$$V_1^\varepsilon(n) = -2\varepsilon \sum_{j=n}^{\infty} V_x(X_n^\varepsilon)[P(\hat{\xi}_{n-1}^\varepsilon,j-n+1,K(X_n^\varepsilon)|X_n^\varepsilon) - P(K(X_n^\varepsilon)|X_n^\varepsilon)] = S^\varepsilon.$$

$V_1^\varepsilon(n)$ is well defined and $O(\varepsilon)$ by (A1). $V_1^\varepsilon(n)$ is introduced in order to average out the noise $\hat{\xi}_{n-1}^\varepsilon$ in (3.2). Using the definition of $\bar{g}(x)$, we have

$$\text{(3.4)} \quad E_{n+1}^\varepsilon V_1^\varepsilon(n+1)-V_1^\varepsilon(n) = -V_x(X_n^\varepsilon)[\varepsilon\bar{g}(X_n^\varepsilon) + (1-2P(\hat{\xi}_{n-1}^\varepsilon,1,K(X_n^\varepsilon)|X_n^\varepsilon))] + T^\varepsilon$$

where

$$-T^\varepsilon = 2\varepsilon \sum_{j=n+1}^{\infty} E_n^\varepsilon\{V_x(X_{n+1}^\varepsilon)[P(\hat{\xi}_n^\varepsilon,j-n,K(X_{n+1}^\varepsilon)|X_{n+1}^\varepsilon) - P(K(X_{n+1}^\varepsilon)|X_{n+1}^\varepsilon)]$$

$$- V_x(X_n^\varepsilon)[P(\hat{\xi}_{n-1}^\varepsilon,j-n+1,K(X_n^\varepsilon)|X_n^\varepsilon) - P(K(X_n^\varepsilon)|X_n^\varepsilon)]\}.$$

We show that $|T^\varepsilon| = O(\varepsilon^2)$.

First, we simplify T^ε. Note that replacing X_{n+1}^ε in $V_x(X_{n+1}^\varepsilon)$ by X_n^ε only alters the sum by $O(\varepsilon^2)$. By making this replacement, writing $2V_x(X_n^\varepsilon) =$

$O(\varepsilon)$ and using the Markov property (3.5)*, we have the equivalent form (3.6):

$$(3.5)\quad P(\hat{\xi}^{\varepsilon}_{n-1},\ j-n+1,\ K(X^{\varepsilon}_n)|X^{\varepsilon}_n) = E^{\varepsilon}_n P(\hat{\xi}^{\varepsilon}_n,\ j-n,\ K(X^{\varepsilon}_n)|X^{\varepsilon}_n)$$

$$(3.6)\quad -T^{\varepsilon} = O(\varepsilon)\sum_{j=n+1}^{\infty} E^{\varepsilon}_n\{[P(\hat{\xi}^{\varepsilon}_n,\ j-n,\ K(X^{\varepsilon}_{n+1})|X^{\varepsilon}_{n+1}) - P(K(X^{\varepsilon}_{n+1})|X^{\varepsilon}_{n+1})]$$

$$- [P(\hat{\xi}^{\varepsilon}_n,\ j-n,\ K(X^{\varepsilon}_n)|X^{\varepsilon}_n) - P(K(X^{\varepsilon}_n)|X_n)]\}.$$

By (A3) and the law of the mean, (3.6) equals (writing $\delta X^{\varepsilon}_n = X^{\varepsilon}_{n+1}-X^{\varepsilon}_n$)

$$O(\varepsilon)E^{\varepsilon}_n(X^{\varepsilon}_{n+1}-X^{\varepsilon}_n)\sum_{j=n+1}^{\infty}\int_0^1 ds[P_x(\hat{\xi}^{\varepsilon}_n,j-n,K(x)|x) - P_x(K(x)|x)]$$

where $x = X^{\varepsilon}_n + s\delta X^{\varepsilon}_n$ is used in evaluating the argument of the integral. But the last expression is $O(\varepsilon^2)$ by (A3) and the fact that $|X^{\varepsilon}_{n+1}-X^{\varepsilon}_n| = O(\varepsilon)$.

Part 2. Define $V^{\varepsilon}(n) = V(X^{\varepsilon}_n) + V^{\varepsilon}_1(n)$ and note that $V^{\varepsilon}_1(n) = O(\varepsilon)$. Summarizing the calculations in Part 1 yields

$$(3.7)\quad E^{\varepsilon}_n V^{\varepsilon}(n+1)-V^{\varepsilon}(n) \le -\varepsilon V_x(X^{\varepsilon}_n)\bar{g}_{\varepsilon}(X^{\varepsilon}_n) + O(\varepsilon^2),$$

By (A2) and the fact that $X^{\varepsilon}_n \in [x_{\ell},x_u]$, there is a $\gamma > 0$ such that

$$(3.8)\quad E^{\varepsilon}_n V^{\varepsilon}(n+1)-V^{\varepsilon}(n) \le -\varepsilon\gamma\ V(X^{\varepsilon}_n) + O(\varepsilon^2).$$

---

*Recall that $\{\hat{\xi}_j(X^{\varepsilon}_n),\ j\ge n,\ \hat{\xi}_{n-1}(X^{\varepsilon}_n)=\hat{\xi}^{\varepsilon}_{n-1}\}$ is a Markov process with initial condition $\hat{\xi}^{\varepsilon}_{n-1}$.

The $V(X_n^\varepsilon)$ on the right side of (3.8) can be replaced by $V^\varepsilon(n)$ without violating the inequality. Consequently, $EV^\varepsilon(n) \le (1-\varepsilon\gamma)EV^\varepsilon(n-1) + O(\varepsilon^2)$, which implies that $EV^\varepsilon(n) = O(\varepsilon)$ for large n; i.e., that there is a $N_\varepsilon < \varepsilon$ such that $V^\varepsilon(n) = O(\varepsilon)$ for $n \ge N_\varepsilon$. Since $V_1^\varepsilon(n) = O(\varepsilon)$, $EV(X_n^\varepsilon) = O(\varepsilon)$ for $n \ge N_\varepsilon$ also, and the theorem is proved. Q.E.D.

<u>Note on the untruncated</u> $\{X_n^\varepsilon\}$ $(x_\ell = -\infty,\ x_u = \infty)$. If we used (1.2) rather than (1.9), then (3.7) would still hold, and we would still have $V_1^\varepsilon(n) = O(\varepsilon)$. But, since (3.8) would not hold, we have not resolved the problem of showing that (3.7) implies the existence of $\{N_\varepsilon\}$ such that $\{U_n^\varepsilon,\ n \ge N_\varepsilon,\ \text{small } \varepsilon\}$ is tight.

IV. The Limit Theorem

Define

$$\sigma^2 = 1 + 2\sum_{j=1}^{\infty} E^\theta \operatorname{sign} \xi_0(\theta) \operatorname{sign} \xi_j(\theta).$$

<u>Theorem 3</u>. <u>Let</u> $\{n_\varepsilon\}$ <u>be any sequence such that</u> $n_\varepsilon \ge N_\varepsilon$. <u>Under</u> (A1) - (A6), $\{U^\varepsilon(\cdot),\ \varepsilon$ <u>small</u>, $U^\varepsilon(0) = U_{n_\varepsilon}^\varepsilon\}$ <u>converges weakly to the diffusion</u> $U(\cdot)$ <u>defined by</u>

$$dU = -\bar{g}_x(\theta)U dt + \sigma dB. \tag{4.1}$$

<u>Let</u> $n_\varepsilon - N_\varepsilon = q_\varepsilon$. <u>If</u> $\varepsilon q_\varepsilon \to \infty$ <u>as</u> $\varepsilon \to 0$, <u>then the weak limit is the stationary solution of</u> (4.1).

Proof. Part 1. The main work of the proof is in constructing the f^ε and verifying (2.3) - (2.5), where the operator A is the differential generator of (4.1). Then if $\{U^\varepsilon(\cdot),\ \varepsilon \text{ small}\}$ is tight (where $U^\varepsilon(0) = U^\varepsilon_{n_\varepsilon}$), the first conclusion will follow. The tightness proof will follow readily from Theorem 2 of [7] or [11]. The assertion concerning stationarity is not hard to get, but the proof is omitted owing to lack of space. Since $\{U^\varepsilon_{n_\varepsilon}\}$ is tight, any subsequence contains a further subsequence which converges weakly (i.e., in distribution). All limits will be of the form (4.1) and can differ only in the initial condition U(0). Also, any tight sequence $\{U^\varepsilon(0)\}$ contains a subsequence which converges weakly to some random variable. Thus, in view of Theorem 2, we can, without loss of generality, assume that $\{U^\varepsilon_{n_\varepsilon}\}$ converges weakly to some random variable U(0). Noting that $|U^{\varepsilon,N}_n| \le N$, we have $X^{\varepsilon,N}_n \in [x_\ell+\varepsilon, x_u-\varepsilon]$, and $\theta \in [x_\ell+\varepsilon, x_u-\varepsilon]$ for small ε. Suppose that ε is small enough for the last sentence to hold. Then $a_\varepsilon(X^{\varepsilon,N}_n) = b_\varepsilon(X^{\varepsilon,N}_n) = \varepsilon$. So the $\{U^{\varepsilon,N}_n\}$ for the truncated problem (1.9) - (1.10) are actually given by (2.2) for small ε, and we use (2.2) henceforth.

Part 2. Fix $f(\cdot,\cdot) \in \hat{\mathscr{L}}^{2,3}_0 = \mathscr{D}$, which is dense in $\hat{\mathscr{L}}_0$, as required by Theorem 1. We drop the superscripts N for the sake of notational simplicity, but we are actually working with $(U^{\varepsilon,N}_n, X^{\varepsilon,N}_n, U^{\varepsilon,N}(\cdot), X^{\varepsilon,N}(\cdot), E^{\varepsilon,N}_n)$ and not with $(U^\varepsilon_n,\ldots)$ henceforth. We have

$$(4.2)\quad E^\varepsilon_n f(U^\varepsilon_{n+1}, n\varepsilon+\varepsilon) - f(U^\varepsilon_n, n\varepsilon) = \varepsilon f_t(U^\varepsilon_n, n\varepsilon) + o(\varepsilon) + f_u(U^\varepsilon_n, n\varepsilon) E^\varepsilon_n (U^\varepsilon_{n+1} - U^\varepsilon_n)$$
$$+ \frac{\varepsilon}{2} f_{uu}(U^\varepsilon_n, n\varepsilon) b^2_N(U^\varepsilon_n),$$

where

$$E^\varepsilon_n (U^\varepsilon_{n+1} - U^\varepsilon_n) = \sqrt{\varepsilon}\, b_N(U^\varepsilon_n)[1 - 2P(\hat{\xi}^\varepsilon_{n-1}, 1, K(\theta+\sqrt{\varepsilon}\, U^\varepsilon_n) | \theta+\sqrt{\varepsilon}\, U^\varepsilon_n)].$$

Actually, since $U^{\varepsilon,N}(0) = U_{n_\varepsilon}$ is the initial condition for $U^{\varepsilon,N}(\cdot)$, all indices n should be $n_\varepsilon+n$ and the time argument $n\varepsilon$ should be $n\varepsilon - n_\varepsilon \varepsilon$. But we "shift" to the origin for notational simplicity.

The third term of (4.2) must be "averaged out", and this requirement determines the form of the f^ε. We will look for $f^\varepsilon(\cdot)$ of the form $f^\varepsilon(n\varepsilon) = f(U_n^\varepsilon, n\varepsilon) + f_0^\varepsilon(n\varepsilon) + f_1^\varepsilon(n\varepsilon)$. For our choices, the f_i^ε will all be $O(\sqrt{\varepsilon})$ and (2.3) - (2.5) will hold.

Define

$$f_0^\varepsilon(n\varepsilon) = -\sqrt{\varepsilon}\ 2b_N(U_n^\varepsilon) f_u(U_n^\varepsilon, n\varepsilon) \sum_{j=n}^{\infty} [P(\hat{\xi}_{n-1}^\varepsilon, j-n+1, K(X_n^\varepsilon) | X_n^\varepsilon) - P(K(X_n^\varepsilon) | X_n^\varepsilon)].$$

By (A1), $f_0^\varepsilon(n\varepsilon)$ is well defined. We have

$$\text{(4.3)}\quad E_{n+1}^\varepsilon f_0^\varepsilon(n\varepsilon+\varepsilon) - f_0^\varepsilon(n\varepsilon) = o(\varepsilon) - \text{third term of (4.2)}$$

$$+ \sqrt{\varepsilon}\ b_N(U_n^\varepsilon) f_u(U_n^\varepsilon, n\varepsilon)(1 - 2P(K(X_n^\varepsilon) | X_n^\varepsilon))$$

$$- 2\sqrt{\varepsilon} \sum_{j=n+1}^{\infty} E_n^\varepsilon b_N(U_{n+1}^\varepsilon) f_u(U_{n+1}^\varepsilon, n\varepsilon) [P(\hat{\xi}_n^\varepsilon, j-n, K(X_{n+1}^\varepsilon) | X_{n+1}^\varepsilon) - P(K(X_{n+1}^\varepsilon) | X_{n+1}^\varepsilon)] +$$

$$+ 2\sqrt{\varepsilon} \sum_{j=n+1}^{\infty} E_n^\varepsilon b_N(U_n^\varepsilon) f_u(U_n^\varepsilon, n\varepsilon) [P(\hat{\xi}_{n-1}^\varepsilon, j-n+1, K(X_n^\varepsilon) | X_n^\varepsilon) - P(K(X_n^\varepsilon) | X_n^\varepsilon)].$$

Let T_1^ε, T_2^ε denote the last two terms. Noting that $2P(K(x)|x) - 1 = \bar{g}(x)$, we can write the third term on the right of (4.3) as

$$\text{(4.4)}\quad -\sqrt{\varepsilon}\ b_N(U_n^\varepsilon) f_u(U_n^\varepsilon, n\varepsilon) \bar{g}(X_n^\varepsilon) = -\varepsilon b_N(U_n^\varepsilon) f_u(U_n^\varepsilon, n\varepsilon) \bar{g}_x(\theta) U_n^\varepsilon + o(\varepsilon).$$

Next, we simplify the T_i^ε. Write $b_N(U_{n+1}^\varepsilon) f_u(U_{n+1}^\varepsilon, n\varepsilon) = b_N(U_n^\varepsilon) f_u(U_n^\varepsilon, n\varepsilon) + (b_N(U_n^\varepsilon) f_u(U_n^\varepsilon, n\varepsilon))_u (U_{n+1}^\varepsilon - U_n^\varepsilon) + O(\varepsilon)$, and split T_1^ε into the three corresponding sums: $T_1^\varepsilon = T_{11}^\varepsilon + T_{12}^\varepsilon + T_{13}^\varepsilon$. Now $T_{13}^\varepsilon = O(\varepsilon^{3/2}) = o(\varepsilon)$. Noting (3.5), combine $T_{11}^\varepsilon + T_2^\varepsilon$ and use a differentiability argument such as used below (3.6), together with the fact that $|X_{n+1}^{\varepsilon}-X_n^{\varepsilon}| \le \varepsilon$, to get that $T_{11}^{\varepsilon} + T_2^{\varepsilon} = o(\varepsilon)$. The above simplifications of (4.3) yield

$$(4.5)\quad E_{n+1}^{\varepsilon} f_0^{\varepsilon}(n\varepsilon+\varepsilon) - f_0^{\varepsilon}(n\ ) = -\varepsilon b_N(U_n^{\varepsilon}) f_u(U_n^{\varepsilon}, n\varepsilon)\bar{g}_x(\theta) U_n^{\varepsilon} + o(\varepsilon) - \text{third term of (4.2)}$$

$$- 2\sqrt{\varepsilon}\cdot \sum_{j=n+1}^{\infty} (b_N(U_n^{\varepsilon}) f_u(U_n^{\varepsilon}, n\varepsilon))_u E_n^{\varepsilon}(U_{n+1}^{\varepsilon} - U_n^{\varepsilon}) [P(\hat{\xi}_n^{\varepsilon}, j-n, K(X_{n+1}^{\varepsilon}) | X_{n+1}^{\varepsilon})$$

$$- P(K(X_{n+1}^{\varepsilon}) | X_{n+1}^{\varepsilon})].$$

By another differentiability argument it can be shown that if the X_{n+1}^{ε} in the P terms of (4.5) are replaced by θ, then the sum changes only by $o(\varepsilon)$. Make this replacement and note that $P(K(\theta)|\theta) = 1/2$. The revised sum in (4.5) is

$$(4.6)\quad \varepsilon(b_N(U_n^{\varepsilon}) f_u(U_n^{\varepsilon}, n\varepsilon))_u \sum_{j=n+1}^{\infty} 2E_n^{\varepsilon} \operatorname{sign}[k(X_n^{\varepsilon}) + \xi_n^{\varepsilon}][P(\hat{\xi}_n^{\varepsilon}, j-n, K(\theta)|\theta) - \tfrac{1}{2}] + o(\varepsilon).$$

Next, it is not hard to see that by dropping the $k(X_n^{\varepsilon})$ term in (4.6), the sum changes only by $o(\varepsilon)$. To see this, use (A1) and the fact that $|U_n^{\varepsilon}| \le N+1$ and also (implied by (A5)) $E_n^{\varepsilon}\ |\operatorname{sign}[k(\theta+\sqrt{\varepsilon}U_n^{\varepsilon}) + \xi_n^{\varepsilon}] - \operatorname{sign}(\xi_n^{\varepsilon})| \to 0$ as $\varepsilon \to 0$, uniformly in $\hat{\xi}_{n-1}^{\varepsilon}$ and in $X_n^{\varepsilon} \in [x_\ell, x_u]$. <u>Let us drop this</u> $k(X_n^{\varepsilon})$ <u>term</u>. In addition, (A4) and (A1) imply that if the E_n^{ε} in (4.6) were replaced by E_n^{θ} (and $\hat{\xi}_n^{\varepsilon}$, ξ_n^{ε} by $\hat{\xi}_n(\theta)$, $\xi_n(\theta)$), then the sum would change by $o(\varepsilon)$ only. This last assertion follows from the observation that, after the $k(X_n^{\varepsilon})$ is dropped, the jth summand is actually $E_n^{\varepsilon}(\operatorname{sign} \xi_n^{\varepsilon}) E_{n+1}^{\theta} \operatorname{sign} \xi_j(\theta)$. Then note that

$$E_n^{\theta} \operatorname{sign} \xi_n(\theta) \sum_{j=n+1}^{\infty} E_{n+1}^{\theta} \operatorname{sign} \xi_j(\theta) - E_n^{\varepsilon} \operatorname{sign} \xi_n^{\varepsilon} \sum_{j=n+1}^{\infty} E_{n+1}^{\theta} \operatorname{sign} \xi_j(\theta)$$

$$\equiv E_n^{\theta} F(\hat{\xi}_n(\theta)) - E_n^{\varepsilon} F(\hat{\xi}_n^{\varepsilon})$$

where by (A1), $F(\cdot)$ is uniformly bounded in n, ε. Now use (A4). Then we can write

$$(4.7) \qquad (4.6) = o(\varepsilon) + \varepsilon(b_N(U_n^\varepsilon)f_u(U_n^\varepsilon,n\varepsilon))_u \sum_{j=n+1}^{\infty} E_n^\theta \text{ sign } \xi_n(\theta) \text{ sign } \xi_j(\theta).$$

Part 3. We now "average out" the sum in (4.7). Define

$$f_1^\varepsilon(n\varepsilon) = \varepsilon(b_N(U_N^\varepsilon)f_u(U_n^\varepsilon,n\varepsilon))_u \sum_{j=n}^{\infty}\sum_{\ell=j+1}^{\infty}[E_n^\theta \text{ sign } \xi_j(\theta) \text{ sign } \xi_\ell(\theta)$$
$$- E^\theta \text{ sign } \xi_j(\theta) \text{ sign } \xi_\ell(\theta)].$$

The sum is well defined by (A6). By an argument similar to that used in Part 2, we can show that

$$E_n^\varepsilon f_1^\varepsilon(n\varepsilon+\varepsilon)-f_1^\varepsilon(n\varepsilon) = -(4.7) + o(\varepsilon)$$
$$+ \varepsilon(b_N(U_n^\varepsilon)f_u(U_n^\varepsilon,n\varepsilon))_u \sum_{j=1}^{\infty} E^\theta \text{ sign } \xi_0(\theta) \text{ sign } \xi_j(\theta).$$

Finally, summing up the calculations and cancelling terms whenever possible yields

$$|f_0^\varepsilon(n\varepsilon)| + |f_1^\varepsilon(n\varepsilon)| = O(\sqrt{\varepsilon}),$$

$$E_n^\varepsilon f^\varepsilon(n\varepsilon+\varepsilon)-f^\varepsilon(n\varepsilon) = o(\varepsilon) + \varepsilon f_t(U_n^\varepsilon,n\varepsilon) - \varepsilon b_N(U_n^\varepsilon)f_u(U_n^\varepsilon,n\varepsilon)\bar{g}_x(\theta)U_n^\varepsilon$$
$$+ \frac{\varepsilon}{2}f_{uu}(U_n^\varepsilon,n\varepsilon)b_N^2(U_n^\varepsilon) + \varepsilon(b_N(U_n^\varepsilon)f_u(U_n^\varepsilon,n\varepsilon))_u \sum_{j=1}^{\infty} E^\theta \text{ sign } \xi_0(\theta) \text{ sign } \xi_j(\theta).$$

Thus (2.3) - (2.5) hold for an operator A^N which equals the operator A of (4.1) in each S_N.

Tightness of $\{U^{\varepsilon,N}(\cdot)$, small $\varepsilon\}$. The tightness proof of Theorem 2 of [7] or [11] requires the construction of $f^{\varepsilon}(n\varepsilon)$ similar to those used here - for each N. But since our $f_i^{\varepsilon}(n\varepsilon)$ are $O(\sqrt{\varepsilon})$, the conditions of the cited theorem hold and we get the tightness immediately in our case. Q.E.D.

Remark. In more general cases the $f^{\varepsilon,N}$ are chosen in a similar way: $f^{\varepsilon,N} = f$ + small perturbation. First we obtain an expansion (up to $o(\varepsilon)$) of $E_n^{\varepsilon,N} f(Y_{n+1}^{\varepsilon,N}, n\varepsilon+\varepsilon) - f(Y_n^{\varepsilon,N}, \varepsilon n)$. Then check which terms in the expansion need to be averaged out - or replaced by an "average value". These will be the terms which do not depend solely on $Y_n^{\varepsilon,N}, \varepsilon n$. Then the sum $f_0^{\varepsilon,N}$ is introduced (centered about a "mean value" - which is the averaged replacement for the undesirable term). Continue as in the proof, building up the operator A^N step by step.

References

1. H.J. Kushner, D.S. Clark, Stochastic approximation methods for constrained and unconstrained systems, Springer, 1978.

2. L. Ljung, "Analysis of recursive stochastic algorithms", IEEE Trans. on Automatic Control, AC-22, 1977, pp. 551-575.

3. H.J. Kushner, Hai Huang, "Rates of convergence for stochastic approximations with constant coefficients", SIAM J. on Control and Optimization, 17, 1979, pp. 607-617.

4. H.J. Kushner, Hai Huang, "Asymptotic properties of stochastic approximations with constant coefficients", to appear, SIAM J. on Control and Optimization.

5. A. Benveniste, M. Goursat, G. Ruget, "Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms, with applications to data communication", preprint IRIA, 1979.

6. A. Benveniste, M. Goursat, G. Ruget, "Analysis of a non-minimum phase system: blind adjustment of a linear equalizer in data communications", preprint, IRIA, 1979.

7. H.J. Kushner, Hai Huang, "On the weak convergence of a sequence of general stochastic difference equations to a diffusion", submitted to SIAM J. on Applied Math.

8. H.J. Kushner, Hai Huang, "Asymptotic properties of learning and adaptation algorithms", in preparation.

9. D.W. Strook, S.R.S. Varadhan, Multidimensional diffusion processes, Springer, New York, 1979.

10. P. Billinglsey, Convergence of probability measures, Wiley, New York, 1968.

11. H.J. Kushner, "A martingale method for the convergence of a sequence of processes to a jump-diffusion process", to appear, Z. Wahrscheinlichkeitstheorie.